PROCEEDINGS *of the* FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

Held at the Statistical Laboratory
University of California
June 21–July 18, 1965
and
December 27, 1965–January 7, 1966

with the support of
University of California
National Science Foundation
National Institutes of Health
Air Force Office of Scientific Research
Army Research Office
Office of Naval Research

VOLUME IV

BIOLOGY AND PROBLEMS OF HEALTH

EDITED BY LUCIEN M. LE CAM
AND JERZY NEYMAN

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1967

A DECISION THEORETICAL COMPARISON OF THREE PROCEDURES OF SCREENING FOR A SINGLE DISEASE

CHARLES D. FLAGLE
THE JOHNS HOPKINS UNIVERSITY

1. Introduction

We should begin by making a distinction between the terms "screening" and "diagnosis." The purpose of screening is to select from an apparently healthy population those who display sufficient probability of an illness to warrant referral for diagnosis. As defined by the National Conference on Chronic Diseases [47] (cited in [16]): "Screening is the presumptive identification of unrecognized disease or defect by the applications of tests, examinations or other procedures which can be applied rapidly." Thus, screening is not a decision about therapy, but a method of case finding and a step toward diagnosis. Though the emphasis in this paper is on analysis for a single disease, this may be done as a part of a multiple or multiphasic screening program. In fact, the point of interest is how, from a multitude of data, relevant information may be recognized and combined to increase the precision of screening. We shall examine here three ways of using screening data: the single test, with positive or negative indication; the profile, an array of estimates of levels for each of a set of relevant factors; and the index, a single composite of weighted factors.

As a first stage in the medical care process, screening has evoked controversy over safety, effectiveness, and economy; and the purpose here is to examine some of the issues in a decision theory context. This paper is a continuation of two earlier discussions; that of Churchman [14] in his treatment of values, and Chiang, Hodges, and Yerushalmy [13] in the treatment of statistics.

The literature on screening is large, though scattered. References [7], [8], [16], [34], [47], [49], and [54], describe the problem from the medical historical point of view. Blumberg [4], Scheff [51], and Thorner [56] have introduced some of the value and decision theory considerations pursued here. Federer [20] has compiled an extensive bibliography on the generic problem of screening.

The decision process of interest in screening for disease is one of policy making. A large number of persons are to be examined and a wide range of manifestations is expected. The problem is to decide beforehand what action is to be taken over sets of manifestations, taking into account such factors as prevalence of

This work is supported in part by USPHS Research Grant HM-00279.

the disease screened for, the cost of screening, the costs of missed cases (false negatives), and unnecessary referrals (false positives), and the constraints of the capacity of the medical care system to handle the referrals from the screening process. We have the extensive form of the decision problem of determining strategy, the beforehand choice of actions to be taken in response to any of the possible test results when they are encountered. In the diseases of interest here, there are not likely to be tests that are pathognomonic, that is, capable of judging precisely the presence or absence of the disease. Most tests behave as the name screening implies; as the "mesh" is tightened to catch more of the positive cases, it catches an increasing number of negative subjects as well. An increase in sensitivity is often bought at the expense of a loss of specificity, and the decision to refer a case or not must be one taken under risk or uncertainty. The approach proposed is to determine from the available range of screening observations and tests, one, or the level of one, or a group, which minimizes the sum of costs of missed cases and unnecessary referrals; then to compare this with such alternative strategies as (a) screening and referring no one, or (b) bypassing screening directly into the referral process.

In this paper, the problem is examined from the point of view of the effect of disease on society at large as well as on the individuals afflicted and the system of health services which must care for them. The value considerations are complex ones, for with several segments of society involved, the process is one of group decision. In the past it has been possible to avoid formal treatments of the value problem, for either severely constrained resources have dominated decisions, or ignorance of statistical properties of the disease and its detection would have prevented the application of decision theory, even if meaningful measures of values and costs were available. Both obstacles are giving way here and there and it becomes ultimately necessary to treat the value problem. One aim of this study is to attempt to estimate relevant values, utilities or losses in the context of the very decision procedures in which they are needed.

A traditional approach to evaluation of screening is to determine whether screening is justified at all. It compares the yield, that is, the number of previously undetected cases discovered, to the cost of the screening program itself plus the cost of follow up examination. Another figure of merit used is the confirmation ratio, the proportion of referrals confirmed as true positive cases [9], for if prevalence of true disease is low and specificity of screening is not total, the follow up resources may be deluged with false positives. The occurrence of false negatives, the missed cases, has been considered an inherent danger of screening; it is a basis for the argument that a false negative indication may engender unjustified confidence. On the other hand, it is argued by proponents of screening that the risk of biasing missed cases against further examination is justified by the benefits to those detected. As cited in [16] ". . . our satisfaction in the number detected far outweighs our grief over those missed."

Evaluation of screening processes from a decision theoretic point of view accepts screening *pro tem* as a potentially admissible strategy and introduces

the cost of missed cases. However, the current state of knowledge does not yet permit a decision theoretic formulation in the case of some important diseases. The widely used procedure in screening for glaucoma, for example, is measurement of intraocular pressure, a variable associated with the disease [7], and referral of those whose pressure exceeds two—or perhaps three—standard deviations from the population mean. Many existing screening procedures and the criteria for evaluating them have this characteristic of limiting the referral cases. Packer, Deutsch, Deweese, Kashgarian, and Lewis [42] have studied the distribution of the variable in a large sample, but unfortunately, although some follow up has been made in low pressure groups, they do not have sufficient data to establish the form of the conditional distribution of pressure given existence of the disease. Sensitivity appears to be low, and one is motivated to find better screening measures as well as a rationale for their use in decision.

There are many single tests for which estimates of sensitivity and specificity are available. Since these form the basis for most current practice we shall examine first some of the implications of the single test.

2. The single test

Single tests form the basis for most current procedures, whether used alone in a single disease screening or as part of a multiphasic screening program. Most familiar are, for example, chest X-ray for tuberculosis detection [18], [51], tonometry for glaucoma [6], [26], [42], [44], blood or urine analysis for diabetes [33], [48], and several tests for heart disease [32], [50]. The accuracy of test procedures is treated in [25], and [39]. The results of tests may be a dichotomous positive or negative indication, such as the presence or absence of bacilli in a smear [22] or it may be a scale reading of a continuous variable, such as intraocular pressure. The interesting problem in the latter case is to select a screening level or levels, thresholds beyond which a change in action should be made. A point to be kept in mind is that the test may be self-administered, or administered by a technician. Subtleties of medical judgment cannot be assumed at the screening stage.

The problem lends itself to expression and solution in the form of a two dimensional game against nature. We consider only two states of nature; the positive state, θ_1: "ought to be referred for further examination" and the negative state θ_2: "need not be referred now." (In some instances, a third state such as the critical one "ought to be treated immediately" is recognized.) The states of nature have been operationally defined above in terms of the available actions, referral a_1, or, dismissal a_2. The meaning of the test indications, positive x_1, or negative x_2, is also implied, and a key element in the problem is the precision of the test, expressed conveniently as sensitivity, the proportion of true positives giving a positive indication $p(x_1|\theta_1)$, and specificity, the proportion of true negatives giving a negative indication $p(x_2|\theta_2)$. If it is possible to express losses or regrets associated with each action for each state of nature $L(a, \theta)$ and

to identify the gamut of strategies, then it is possible to compute an expected loss for each strategy for each state of nature

$$L(\theta, s) = \sum_a \sum_x p(x|\theta)p(a|x, s)L(a|\theta), \tag{2.1}$$

where $p(a|x, s)$ is 1 if in strategy s, x calls for action a, and 0 if it does not.

If *a priori* probability P, that is, prevalence of the disease in the screened population is known, an expected loss can be computed for each strategy,

$$E[L(\theta, s)] = PL(\theta_1, s) + (1 - P)L(\theta_2, s). \tag{2.2}$$

Three significant strategies need to be examined. First there is the strategy, call it s_1, of nonscreening—or not responding to screening, whatever its indication. This incurs for true positives the loss $L(a_2, \theta_1)$, the loss of not detecting a true positive case, but it places no burden of false positives $L(a_1, \theta_2)$ on the referral system. It is the strategy of doing nothing to seek out cases, and as a practical matter, is often defended on the grounds of inadequacy or inherent ineffectiveness of the system to handle referrals. It may be defensible also if lack of specificity in screening and low prevalence incur referral costs greater than costs of missed cases.

Another strategy s_2 would bypass a screening procedure and send everyone to the referral process. Here the cost of unnecessary referrals $L(a_1, \theta_2)$, is incurred with certainty but in return, false negatives are avoided. Note that a mixed strategy of s_1 and s_2 is possible by selecting some proportion of the population at random for referral without screening.

Finally there is strategy s_3 of responding to a screening procedure, referring those judged positive for diagnosis and dismissing those with negative indication. There is of course a fourth strategy, the perverse one of responding in a way contrary to the screening indications, but this need not be considered.

Hopefully, s_3 is the best strategy but whether it is or not depends upon prevalence of the suspected disease as well as the precision of screening. The decision problem can be illustrated graphically in the manner of Chernoff and Moses [11]. In figure 1 we consider only two kinds of loss, those associated with errors, letting R_1 be the regret of missing a case, and R_2 be the regret of an unnecessary referral. Thus, strategy s_1 incurs no regret for the undiseased and s_2 none for the diseased. A straight line in figure 1 connecting the intercepts R_1 and R_2 gives the regrets for mixed s_1 and s_2 strategies. The screening strategy is optimal if on the convex set containing the other strategies, it is the first to be supported by a parametrically increasing family of lines of constant regret. The dotted curve in figure 1 represents a screening test in which a variable x, such as intraocular pressure or blood sugar, is measured and some level of the measurement, x_1, must be chosen as the threshold for a positive indication. In this case s_1 and s_2 are special cases of s_3, the extremes of the possible assignment of x_1.

Using this diagram, we can identify and explain some of the problems and controversy that have attended the consideration of screening policy. In figure 2 the diagram is repeated to show some examples of screening results. The origin

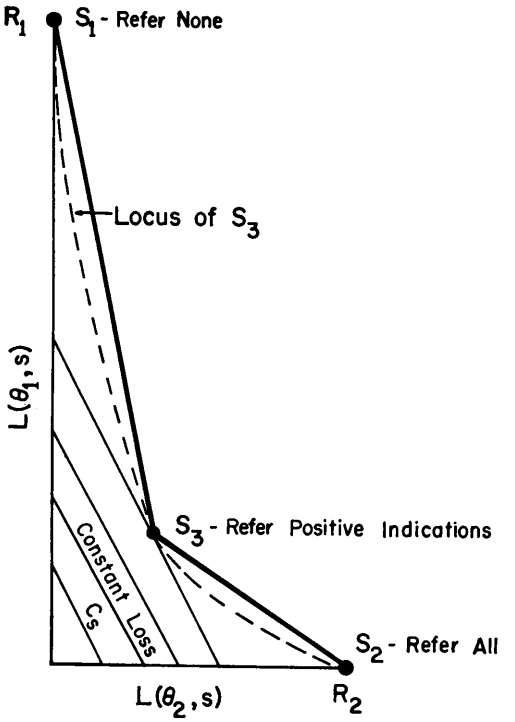

FIGURE 1

Graphical representation of screening decision problem. Dashed line shows locus of S_3 if x_1 is an arbitrary level of a continuous variable.

represents perfect screening, but with loss of sensitivity and specificity, the points radiate toward the line of mixed nonscreening strategies. The closer they get to this line, the narrower the range of prevalence for which screening is optimal. The narrower this range, the more sensitive the solution becomes to all the estimates—of prevalence, sensitivity, specificity, and losses. Diabolically, the usual low prevalence of the screened disease is accompanied by a high ratio of losses or regrets of untreated cases to unnecessary referrals R_1/R_2. The supporting lines of constant cost and the line of mixed nonscreening policies are of nearly similar negative slopes and the choice between these extreme strategies is either indifferent or sensitive to the estimates of prevalence and regrets. Thus, the real world controversy over the value of screening is understandable. Screening is a means of resolving the controversy by providing a strategy which may dominate the others. Obviously, the greater the precision of screening, the wider the range of prevalences for which it is optimal and the less sensitive the solution is to errors in estimate of any of the significant variables. Analyses of the sensitivity of solutions to estimates of the prior distribution and the losses have

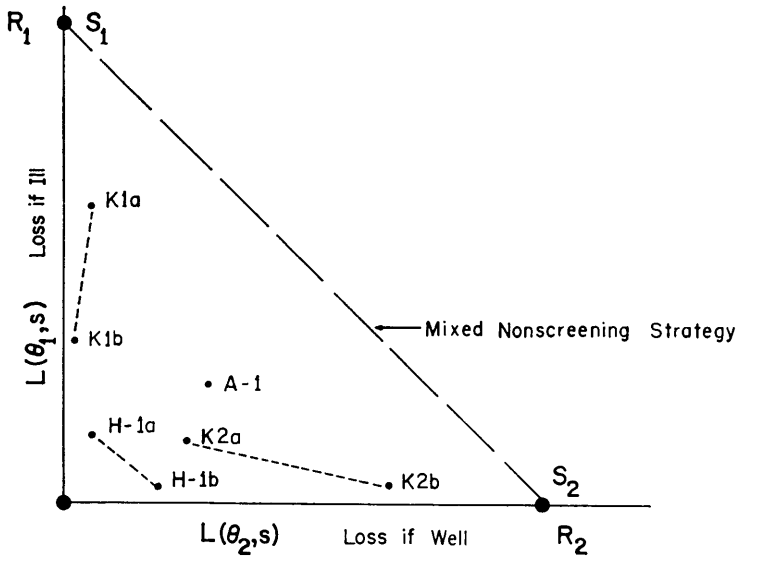

FIGURE 2

Result of some screening studies.
A-1 [1], General Health (CM1);
H-1a [28], General Health 85.5, 93.2;
H-1b [28], General Health (with exam) 98.1, 76.9;
K-1a [32], Diabetes (urinalysis) 38.1, 92.9;
K-1b [32], Diabetes (blood sugar) 66.7, 97.8;
K-2a [33], Heart disease (blood press.) 86.3, 74.0;
K-2b [33], Heart disease (total battery) 96.6, 31.9.

been made by Pierce [46] and Bovey [5], respectively. One is strongly motivated to avoid these concerns by improving screening precision. One approach is to draw upon more information than that contained in the single test. Certain demographic data as well as additional tests are sources of additional information that may reduce uncertainty, but at the cost of analytical complication. We distinguish two approaches to the handling of multiple measurements. Under the term *profile* are those schemes that maintain the separateness of elements of information, under the term *index* are the procedures for pattern recognition that combine elements of information additively.

3. Profile

Multiple or multiphasic screening offers a means of bringing additional factors to bear on a single disease decision. Although the intent of the total program may be to screen for several diseases, and as noted by Breslow [8], this is more efficient than a set of single disease programs, the decision for any particular

disease may draw on all the meaningful data. Usually the levels of each of a set of relevant factors is estimated for each subject, as for example the classification of psychiatric patients by Overall and Gorham [41], of 16 factors and three levels. The levels may be dichotomous, for example, yes or no answers to questions, or continuous variables divided into cohorts. The need for discreteness is to permit a finite number of combinations of factors. Each subject then presents a pattern or profile; a population profiled in n factors at m levels of each has a possible m^n profile. An example of this approach is given by Collen, Rubin, Neyman, Dantzig, Baer, and Seigelaub [15], in screening for asthma. Yes and no answers to six questions yield 64 possible profiles. If a sufficiently large population is screened simultaneously and all are then confirmed positive or negative by a reliable diagnosis, it is possible as shown in [15] to apply Neyman's method [38] to compute for each profile a likelihood ratio, defined as the ratio of the fraction of confirmed positives in the profile to confirmed negatives in the profile.

This approach has statistical and computational problems, due chiefly to the rapid increase in the number of potential profiles as factors, and levels within factors are added. The introduction of computers has facilitated work with profiles; Kleinmuntz [31] reports cutting time by a factor of 100 in grouping patients into profiles. Parker [43] has developed a computational procedure for recognizing relevant combinations of indications in a many factor problem with potentially a very large number of profiles.

The analytical treatment of profiles may proceed in several ways. They may be arrayed in order of their likelihood ratio and each profile evaluated for its contribution to the cumulative sensitivity and specificity. Then by the method used for the single test for determining screening level—the optimal combination of sensitivity and specificity—the profiles may be divided into negative and positive groups.

It is somewhat simpler to determine from examination of the observation for each profile whether the losses for accepting the profile as positive are greater than those for accepting it as negative. The decision rule, where $L(\text{pos})$ is the incremental loss caused by declaring a profile positive, is to minimize, where x_i is the ith profile,

$$\begin{aligned} L(\text{pos}) &= p(x_i|\theta_2)(1 - P)R_2, \\ L(\text{neg}) &= p(x_i|\theta_1)PR_1. \end{aligned} \tag{3.1}$$

If profiles are arrayed in descending order of likelihood ratio, the positive classification will be chosen until the further decline in $L(\text{pos})$ is first offset by $L(\text{neg})$ at which point the optimum likelihood ratio

$$\frac{p(x_i|\theta_1)}{p(x_i|\theta_2)} = \frac{(1 - P)R_2}{PR_1}. \tag{3.2}$$

Note that the computation of the likelihood ratio on the left involves conditional probabilities only, whereas the criterion or threshold value contains both prevalence and the cost of errors. Both of these may vary in time and place.

Hence, the choice of criterion level becomes a matter of local policy and conditions, while the constituents of the likelihood ratio itself may be universal in character.

In some cases the profile information may be used to supplement the indications of a test. One way to do this is to correlate profile factors such as age or sex with *a priori* probability of a subject's being positive. This is meaningful if the test involves choice of a screening level, for as shown earlier, the optimal level is a function of the *a priori* probability of the disease.

4. The index

Anyone who works with profiles becomes aware of the problems of identifying profile patterns, limiting the number of levels and factors to keep the number of profiles in reasonable bounds, and performing analytical operations. There is a temptation to seek to construct some single measure, an index, which demonstrates the composite intensity of the factors. It opens the possibility of weighting factors according to their importance and summing them or combining them in some other way. A basic problem as noted by Mainland [37] is that any value of a particular index, determined by summing weighted levels of factors, can be arrived at in a multitude of ways—individuals with widely varying profiles may yield the same index. Thus, much of the profile is obscured.

Nevertheless, though unattractive in the respect mentioned, many indices have been empirically useful. In other kinds of classifications, such as the determination of need for resources where the levels in factors are measured in homogeneous units like dollars or man hours, the index has operational significance. For screening or diagnostic classification, it is compatible with the existence of syndromes and for chronic disease it can be thought of as dynamic measure reflecting the progression of component symptoms in a syndrome.

Brodman [10] has made use of the Cornell Medical Index Questionnaire to compute a total score of weighted symptoms, when weight reflects the significance of a symptom for a particular disease. He set arbitrary thresholds of this score and achieved 44 per cent of confirmed diagnoses with few false positives.

Walton [58], after examining the problems of profiling dispensary patients from a 100 question instrument, devised instead an index to screen patients for a set of disease categories. The scheme is applied to the categories one at a time and hence, can fit the definition of single disease screening as it has been used here. For a given disease category, the results of a self-administered yes or no questionnaire are used to determine an index x_i for the ith disease category, where, using Walton's notation

$$x_i = \sum_{k=1}^{n} w_{ik} s_{ik}, \qquad i = 1, 2, \cdots, n, \tag{4.1}$$

where

$$\begin{aligned} w_{ik} &= \text{weight of the } k\text{th symptom for the } i\text{th questionnaire,} \\ s_{ik} &= 0 \text{ for no, 1 for yes answer to the } k\text{th question.} \end{aligned} \tag{4.2}$$

In Walton's study, each x_i is a component of an n dimensional vector; direction indicates the disease category and length the intensity of the syndrome. The reference for decision is the pattern of end points of vectors for patients confirmed in each disease category. The usefulness of the technique depends upon separability of the clusters for each category and the ability to recognize the vector of each subject screened as belonging in one of the clusters, or clearly belonging in none. Sebestyn [53] has given a thorough treatment of the general problem of pattern recognition and separation of classes. Hopkins [30] uses lung cancer diagnosis to illustrate the use of an index for separation of positive and negative classes.

The rating scale method was used for obtaining symptom weight from a sample of physicians. Eckenrode [18] compares various schemes for obtaining subjective multiple weights and confirms the speed of the rating scale method although less variation among raters is achieved by more time consuming approaches of paired comparisons and rank ordering.

As in the other methods discussed, there remains the problem of choice of a

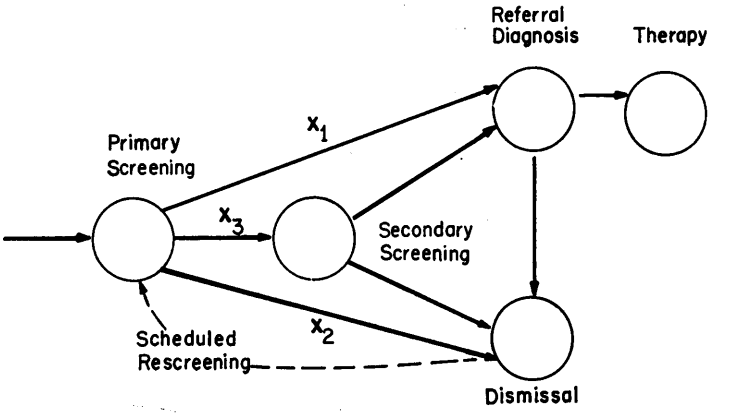

FIGURE 3a

Sequential screening process—patient flow.

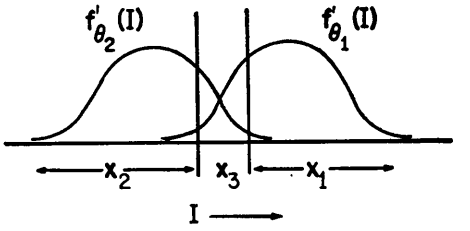

FIGURE 3b

Distribution of indices I, for two states of nature.

decision rule to be applied—how an index is to be used for decision. Abramson, Terespolsky, Brook, and Kark [1] have applied the Cornell Medical Index to a group of patients and computed sensitivity and specificity for some arbitrarily drawn threshold which would permit application of the game against nature previously described. Walton has chosen to present the decision process in a sequential form, to recognize in the index two thresholds—a lower negative screening level, an upper positive screening level, and an intervening region of uncertainty in which the cost and probability of error outweigh the cost of seeking more information. The source of most information in Walton's study was a steering physician, and in fact the purpose of the screening mechanism was to relieve the physician of some of his burden. We have here a potentially two stage model in the form shown in figure 3.

5. Estimation of utilities and losses

Throughout the preceding discussions, we have spoken of values, losses, and regrets as though they were available and meaningful measures subject to mathematical operations. The commensurability of values is an old topic. The classical treatment by Ralph Barton Perry [45] lists three bases for value measurement: intensity of interest, preference, and inclusiveness, where the last term refers to the breadth of the base over which a preference is held. Of these, revealed preference has formed the basis for modern utility theory as promulgated by von Neumann and Morgenstern [57], and indeed, Perry reasons that the measures of interest are contained within preference. However, the problem of inclusiveness is very relevant here; there are many segments of society affected in screening decisions; the patient whose health is at stake, the health services whose resources are to be consumed, the physicians, who traditionally place a high subjective cost on a false negative diagnosis [51], and finally society at large, which suffers from loss of productive capacity of its members through illness. In decision theory terms, we are dealing with group decision, for which, Luce and Raiffa [36] note, a loss table must be arrived at by compromise. The term compromise requires here a broader definition than is usually accorded it; it should embrace the possibilities of integration of conflict in the sense of Follett [24] through admission of changed environment or values.

Much of the experimental work in utility estimation has concerned the development of a utility function for some continuously measurable commodity, such as money. Green [27] has used the standard gamble technique to determine the utility function of some corporate executives for dollars (their own) and rate of return on investment of company funds. Davidson, Siegel, and Suppes [17] have developed experimental procedures for utility estimation. The form of the utility problem in screening procedures is to develop entries for a loss table, reduced in simplest form to the regrets associated with the missed case R_1, and unnecessary referral R_2. The first contains many elements related to undetected and hence perhaps uncontrolled disease: pain and premature death, lost economic

usefulness, possible contagion, lost faith in screening procedure and its participants; R_2, the man hours and facilities wasted in unnecessary referral contains an element of opportunity cost—the value of alternative activity foregone.

A basis for utility estimation almost leaps from the pages, for one can hardly fail to notice that the screening procedure is itself a standard gamble. One strategy is to refer everyone and thus incur with certainty the referral cost. The alternative strategy is screening, with a mixed outcome of some lost cases and some unnecessary referrals. References [22] and [23] give examples of attempts to elicit utilities from decision makers in the context of a specific real screening problem, by diminishing sensitivity of a diagnostic procedure until a threshold of indifference to the procedure is reached. Nunez [40] has attempted to estimate dollar value components of R_1 and R_2 for one disease, glaucoma, and to use them to evaluate strategies.

A realistic backdrop for utility estimation can be developed from the decision process centered on the inclusion of a particular profile or a particular screening level in the positive classification. There is, for example, in [15] an array of profiles from a screening test and follow up for asthma. For each profile, we see the absolute number of confirmed positive and negative cases. Should a particular profile, say one with 5 positive and 33 negative cases, have been included in the positive profile group, had there been no follow up? The question is equivalent to: Does the regret of missing 5 cases exceed the regret of referring 33 unnecessarily? (Apparently it does not in the case at hand, for this profile is not included as positive by the authors, although the profile with the next higher likelihood ratio, having 1 confirmed positive positive and 5 negatives is.)

In order to assess the magnitudes of various components of decision makers' utilities, the questions may be repeated with changes in assumptions. One may assume, for example, in the first questioning that no mechanism exists for regularly repeating screening, thus making a high value of R_2 possible. Introducing a planned repeat of the screening program diminishes the danger to missed cases and the corresponding regret R_2, diminishes to the loss which may occur in the intervening period. Similarly, if the first questioning assumes that all referrals will be made to local resources, the revealed estimate of R_2 may reflect a high opportunity cost attributed to the waste of already constrained resources. A revised assumption that the screening program provides the referral resources, in order to eliminate concern over opportunity costs and capacity constraints, permits the utility estimate to reflect principally the objective cost of the referral examination.

What becomes clear from experiments in this context is that the utilities to be estimated for use in the decision process are *ad hoc*. They depend upon some knowledge of what is to happen next; the time until the next screening and the dynamics of the disease relative to that time interval; the certainty and quality of follow up, the perceived importance of health programs which may compete for follow up resources, the inclusiveness of various decision makers' concern over the missed case. These are brought out in the utility estimation process.

6. Summary and conclusions

An examination of the decision problems in screening leads to several conclusions. First, the more precise, that is sensitive and specific, the screening procedure, the less sensitive is the choice of policy to precision of estimate of other relevant variables such as the prevalence of disease, the losses associated with errors, and the capacity constraints of the follow up system. Hence, there is strong motivation to sharpen precision of screening and the emphasis in this paper is on several approaches, none of which, it appears, is categorically superior to the others.

Second, although the sensitivity, specificity, and accuracy of a screening procedure may be universal and constant, the other variables that influence decision are not. The choice of strategy is dependent upon the losses of missed cases and unnecessary referrals. These losses are affected by the policies and circumstances embodied in the actions taken, for example, whether or not a dismissal is to be followed within a safe period by rescreening. The losses not only differ from place to place, but by virtue of their subjective components they are ephemeral, fortunately so, if their change is required to resolve conflicts in group decision. Hopefully, utility estimation is an evolving process, to be carried out, not remotely, but in the decision making situation.

The problem is dynamic in several senses of that word. The intensity of illness changes in time, so that the regret of false negative dismissal is a function of the interval between screening programs, a problem approached by Lincoln and Weiss [35]. The choice of strategy at the screening stage is dependent upon the actions to be taken at subsequent stages. Thus, screening policies would be formed ideally as part of a total system of management of a disease or a set of diseases.

Most of these conclusions are qualitatively obvious, but it has only been in recent times that one could approach the problem quantitatively in search of optimal policies for which there is hope of implementation. By recent times is meant the time of the computer and the time of bold programs of medical care, for it is clear from the foregoing that the application of rational decision procedures demands knowledge of disease in the apparently healthy population. That is to say, application of decision theory must be accompanied by large and well planned screening and diagnostic studies of the dynamics of important diseases.

REFERENCES

[1] J. H. Abramson, L. Terespolsky, J. G. Brook, and S. C. Kark, "Cornell index as a health measure in epidemiological studies," *Brit. J. Prev. Soc. Med.*, Vol. 19 (1965), pp. 103–110.

[2] J. Berkson, "Utility as a measure on the efficiency of a test," *J. Amer. Stastist. Assoc.*, Vol. 42 (1947), pp. 246–255.

[3] D. Blackwell and M. A. Girschick, *Theory of Games and Statistical Decision*, New York, Wiley, 1954.

[4] M. S. BLUMBERG, "Evaluating health screening procedures," *Operations Research*, Vol. 5 (1957), pp. 351–360.
[5] R. L. BOVEY, "The sensitivity of optimal strategies in finite statistical decision problems to variations in parameters," Ph.D. thesis Johns Hopkins University, 1964.
[6] S. S. BRAV and H. P. KIRBER, "Mass screening for glaucoma," *J. Amer. Med. Assoc.*, Vol. 147 (1951), pp. 1127–1128.
[7] L. BRESLOW, "Chronic disease and disability in adults," Chapter 16 in [49], pp. 507–537.
[8] ———, "Periodic health examination and multiple screening," *Amer. J. Publ. Health*, Vol. 49 (1959), p. 1148.
[9] I. J. BRIGHTMAN, "Follow-up of screening and diagnostic examinations," Chapter 32 in [29], pp. 358–365.
[10] K. BRODMAN, "Diagnostic decisions by machine," *IRE Trans. on Med. Electron.*, Vol. ME-7 (1960), pp. 216–219.
[11] H. CHERNOFF and L. E. MOSES, *Elementary Decision Theory*, New York, Wiley, 1959.
[12] C. L. CHIANG, "On the design of mass medical surveys," *Hum. Biol.*, Vol. 23 (1950), pp. 242–271.
[13] C. L. CHIANG, J. L. HODGES, JR., and J. YERUSHALMY, "Statistical problems in medical diagnosis," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 4, pp. 121–133.
[14] C. W. CHURCHMAN, "Problems of value measurement for a theory of induction and decisions," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 5, pp. 53–59.
[15] M. F. COLLEN, L. RUBIN, J. NEYMAN, G. B. DANTZIG, R. M. BAER, and M. S. SEIGELAUB, "Automated multiphasic screening and diagnosis," *Amer. J. Publ. Health*, Vol. 54 (1964), pp. 744–750.
[16] AMERICAN MEDICAL ASSOCIATION, COUNCIL ON MEDICAL SERVICE, *A Study of Multiple Screening-Descriptive Data on 33 Screening Surveys*, Chicago, 1955.
[17] D. DAVIDSON, S. SIEGEL, and P. SUPPES, "Some experiments and related theory on the measurement of utility and subjective probability," *Stanford Value Theory Report*, SU-AMSL, No. 4, 1955.
[18] R. T. ECKENRODE, "Weighting multiple criteria," *Management Science*, Vol. 12 (1965), pp. 180–192.
[19] P. E. ENTERLINE and H. I. SAUER, "Community wide chest X-ray survey," *Publ. Health Rep., Wash.*, Vol. 66 (1951), pp. 1612–1629.
[20] W. T. FEDERER, "Procedures and designs useful for screening material in selection and allocation" (with bibliography), *Biometrics*, Vol. 19 (1963), pp. 553–587.
[21] C. D. FLAGLE, "A threshold method for utility estimation and resolution of conflict," *Operational Research and the Social Sciences*, London, Tavistock Publications, 1966, pp. 399–408.
[22] C. D. FLAGLE and M. F. LECHAT, "Allocation of medical and associated resources to the control of leprosy," *Management Sciences in the Emerging Countries*, London, Pergamon, 1965.
[23] ———, "Statistical decision theory and the selection of diagnostic and therapeutic procedures in public health," *Actes de la 3me. Conference Internationale de Recherche Operationelle*, London, English Universities Press, 1964.
[24] M. P. FOLLETT, *Dynamic Administration: The Collected Papers of Mary Parker Follett*, (edited by H. C. Metcalf and L. Urwick), New York, Harper, 1942.
[25] L. H. GARLAND, "Studies of the accuracy of diagnostic procedure," *Amer. J. Roentg.*, Vol. 82 (1959), pp. 25–38.
[26] P. GRAHAM and F. C. HOLLOWS, "Sources of variation in tonometry opthalmological transactions," *Trans. Opthalmological Soc., United Kingdom*, 1964, pp. 597–613.

[27] P. J. GREEN, "Measurement of utilities in a large industrial firm," presented at ORSA-TIMS meeting, San Francisco, November, 1961.
[28] J. C. HARVEY, J. W. REED and M. A. THAMER, "Development of a multiphasic screening examination for medical care patients" *J. Chron. Dis.*, Vol. 15 (1962), pp. 827–847.
[29] H. E. HILLEBOE and G. W. LARIMORE (eds.), *Preventive Medicine*, Philadelphia, Saunders, 1965 (2nd ed.).
[30] C. E. HOPKINS, "Discriminant analysis as an aid to medical diagnosis and taxonomy," *J. Indian Med. Profession*, Vol. 11 (1964), pp. 5043–5070.
[31] B. KLEINMUNTZ, "A portrait of the computer as a young clinician," *Behavl. Sci.*, Vol. 8 (1963), pp. 146–149.
[32] A. B. KURLANDER, E. H. HILL, and P. E. ENTERLINE, "An evaluation of some commonly used screening tests for heart disease and hypertension," *J. Chron. Dis.*, Vol. 2 (1955), pp. 427–439.
[33] A. B. KURLANDER, A. P. ISKRANT, and M. E. KENT, "Screening tests for diabetes—a study of specificity and sensitivity," *Diabetes*, Vol. 3 (1954), pp. 213–219.
[34] R. S. LEDLEY and L. B. LUSTED, "Reasoning foundations of medical diagnosis," *Science*, Vol. 130 (1959), pp. 9–21.
[35] T. L. LINCOLN and G. WEISS, "A statistical evaluation of recurrent medical examinations," *Operations Res.*, Vol. 12 (1964), pp. 187–205.
[36] J. D. LUCE and H. RAIFFA, *Games and Decision*, New York, Wiley, 1957.
[37] D. MAINLAND, "Notes from a laboratory of medical statistics," No. 73, 74, 75, 76, 77, Medical Statistics Unit, New York University (mimeographed).
[38] J. NEYMAN, *First Course in Probability and Statistics*, New York, Holt, 1950.
[39] S. NISSEN-MEYER, "Evaluation of screening tests in medical diagnosis," *Biometrics*, Vol. 20 (1964), pp. 730–755.
[40] A. NUNEZ, "Choice of policy in screening for glaucoma," Master's thesis University of California, 1965.
[41] J. E. OVERALL and D. R. GORHAM, "A pattern probability model for the classification of psychiatric patients," *Behavl. Sci.*, Vol. 8 (1963), pp. 108–116.
[42] H. PACKER, A. R. DEUTSCH, M. W. DEWEESE, M. KASHGARIAN, and P. M. LEWIS, "Frequency of glaucoma in three population groups," *J. Amer. Med. Assoc.* Vol. 188 (1964), pp. 123–127.
[43] R. D. PARKER and T. L. LINCOLN, "Medical diagnosis using Bayes theorem," *Health Services Research*, Vol. 2 (1967), pp. 34–45.
[44] E. S. PERKINS, "Glaucoma screening from a public health clinic," *Glaucoma Review*, Vol. 3 (1965), p. 25.
[45] R. B. PERRY, *General Theory of Value*, Cambridge, Harvard University Press, 1954.
[46] D. PIERCE, "A method for determining the sensitivity of a Bayes procedure to the prior distribution," Ph.D. thesis Oklahoma State University, 1965.
[47] "Preventive aspects of chronic disease," *Conference Proceedings, National Conference on Chronic Diseases*, Raleigh, Health Publications Institute, 1952 (cited in [16]).
[48] F. W. REYNOLDS, "Screening methods for diabetes mellitus," Chapter 30 in [29], pp. 349–354.
[49] J. M. ROSENAU and K. F. MAXY, *Preventive Medicine and Public Health* (edited by P. E. Sartwell), New York, Appleton-Century Crofts, 1965 (9th ed.).
[50] D. A. RUTSTEIN, "Screening tests in mass survey, and their use in heart disease case finding," *Circulation*, Vol. 4 (1951), pp. 659–665.
[51] T. J. SCHEFF, "Decision rules, types of errors, and their consequences in medical diagnosis," *Behavl. Sci.*, Vol. 8 (1963), pp. 97–107.
[52] H. H. SCHULTZ, "Screening methods for pulmonary tuberculosis," *Preventive Medecine*, (edited by H. E. Hilleboe and G. W. Larimore), Philadelphia and London, Saunders, 1959.
[53] G. S. SEBESTYN, *Decision Making Processes in Pattern Recognition*, New York, Macmillan, 1962.

[54] W. G. SMILLIE and R. G. HALM, "Inherent inadequacies of multiphase screening," *N. Y. St. J. Med.*, Vol. 52 (1952), pp. 2610–2611.

[55] P. SUPPES, "The role of subjective probability and utility in decision making," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1956, Vol. 5, pp. 61–74.

[56] R. M. THORNER and Q. R. REMEIN, "Principles and procedures in the evaluation of screening for disease," *Public Health Monograph*, No. 67, Washington, U. S. Government Printing Office, 1961.

[57] J. VON NEUMANN and O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton, Princeton University Press, 1953.

[58] W. W. WALTON, "Modern decision theory and medical diagnosis," Ph.D. thesis, Johns Hopkins University, 1964.